算法监控对交通规则遵从意愿的影响：基于解释水平的视角

尹军[1] 肖青露[1] 王诗琪[1] 陈佳锴[1] 许婉秋[1] 范玲霞[2]

([1]宁波大学教师教育学院心理学系暨研究所, 宁波 315211)

([2]宁波行政学院公共管理教研部, 宁波 315211)

[摘　要] 虽然电子警察这类算法监控有诸多优势，但其是否可提升人们对交规的遵从及潜在机制缺乏系统探讨。本研究采用情境测试与田野调查相结合的方法，发现对低解释水平的交通行为(如超速)，相比人类监控，人们对算法监控形成的匹配感(即"感觉合适"的体验)更强，进而提升了对交规的遵守意愿；而对高解释水平的交通行为(如未按规定让行人)，人类监控下人们的匹配感更强，从而有更强的意愿遵守交规。当了解算法出错或失误后，其对低解释水平交通行为遵守意愿的提升作用消失，但人类监控失误时，其对高解释水平交通行为的监控效果依然优于算法监控，匹配感仍然起中介作用。因此，算法监控对交规遵从意愿的提升作用视交通行为解释水平而定，且算法失误削弱其对低解释水平交通行为的监控效果。

[关键词] 算法监控; 人类监控; 交规遵从; 解释水平; 算法失误

[分类号] B842

1 前言

随着城市治理智慧化的推进，越来越多的城市开始引入算法治理(Herath & Mittal, 2022; Oleksy et al., 2023)，其对人们的态度和行为模式也产生了影响(Karimi et al., 2018; Yigitcanlar & Cugurullo, 2020)。作为智慧城市治理的一部分，电子警察这类算法监控常被用于交通违规行为的监测，以提高人们的交通规则(简称"交规")遵从行为。算法监控不仅在一定程度上弥补了道路交通管理在时间和空间上的盲点，且相比交警或辅警进行人类监控，存在诸如人工成本低、监控范围广等优势。虽然算法监控有诸多优势且应用范围广泛，但其是否可提升人们对交规的遵从及其可能机制缺乏系统的实证探讨。对该问题的探讨可揭示算法监控在何种情境下以及如何影响人们的交规遵从行为，以为道路中算法监控的设置提供心理学依据，从而高效利用当前的人工智能算法促进人们对交规的遵从。

---

*浙江省哲学社会科学领军人才培育专项(青年英才培育)，(21QNYC12ZD)资助

*通信作者：尹军，Email: yinjun1@nbu.edu.cn

算法监控的核心是人工智能算法，其可对收集的交通行为进行深度分析，从而实现自动判定交通违规行为(Clarke, 1988; 张笑语, 2021)。算法的智能性和高效性使得其大多时候可超越人类的决策质量，促使算法欣赏(algorithm appreciation; Longoni et al., 2019; Newman et al., 2020)，即相比人类选择和决策，人们更偏好算法。确有研究发现，在具有明确的外部标准时，人们认为算法具有更准确的判断和更有效的检测(Tayarani, 2020)。Laï 等(2020)的调查也发现，外科手术医生对具有智能算法工具的使用持有积极的态度，认为其可提高对患者的照料水平且成本较低。针对视觉数量估计、歌曲流行程度预测以及浪漫关系预测(Logg et al., 2019)以及基于有限的学业信息预测学生的数学成绩(You et al., 2022)，人们更可能采择算法给出的决策建议，表现出算法欣赏。由于算法按照事先设定的规则和程序运行，人们普遍认为算法可忽略被评判个体的独特特征，极大保证了决策过程的客观性和公正性等，从而也会促进算法欣赏(Newman et al., 2020)。如：种族歧视和不平等威胁情境可提升人们对算法决策的欣赏(Bigman et al., 2021; Longoni et al., 2019)。甚者，即使算法做出有歧视性的决策结果(如因“女性”身份而拒绝求职者)，人们采取行动反对这些歧视行为的倾向较弱，且诱发的道德愤怒低于人类的歧视行为(Bigman et al., 2023)，即更能接受算法偏见(algorithmic discrimination)。

算法欣赏使得人们更倾向于选择算法提供的建议或选项，甚至是不道德建议，可促进人们对算法所建议行为的遵从。如 Genda 等(2023)发现电子交通执法(electronic traffic law enforcement)缩短了交通违法行为发生后被处罚的时间，增强了交通法规的威慑作用且减少了该地区的交通违规行为。同时也有研究表明，电子交通执法促进了交警执法和收缴罚款的公平性(Pribadi, 2021)。且以算法为主导的智能交通，有助于减少交通事故中涉及的人为错误(Niestadt et al., 2019)。即使对于更普遍的道德规范，Hertz 等(2019)发现，面对算法实施的行为，人们会像遵从人类行为一样遵从算法。Leib 等(2021)的研究也揭示人类和算法所提供的建议对人们违背道德规范的意愿产生了同样的不利影响。当算法提供决策建议时，即使人们知道该建议具有道德上的缺陷，仍倾向于选择遵从(Hoc & Lemoine, 1998; Krügel et al., 2023)。因此，鉴于算法监控所建议行为为遵从交规，其似乎可提高人们的遵从意愿。

然而，在算法欣赏存在的同时，即使在某些任务中算法决策的表现优于人类，人们仍然偏好选择自己或其他人提供的决策建议，对使用算法持有消极的态度，呈现算法厌恶(algorithm aversion; Dietvorst et al., 2015; Kawaguchi, 2021)。如在个人偏好推荐领域，研究发现相比算法推荐，人们更倾向于选择朋友推荐的书和电影等(Sinha & Swearingen, 2001)。尽管智能算法对学生未来发展的预测准确性高于人类，但人们依然会选择人类的预测作为决策

依据，且当算法与人类出现同样的错误或失误(error)后，人们更容易对算法失去信心(Dietvorst et al., 2015)。进一步研究发现，算法的自动化过程及不透明性让人们难以描述和理解其背后的直接决策过程和原因，且算法失误通常是系统性的且难以修复的，人们认为算法无法像人类那样从错误中学习并改进，故对算法决策质量持怀疑态度且对算法失误容忍度更低(Mahmud et al., 2022; Cadario et al., 2021)。但人们对人类决策有更高的容错度，因为相比固定程序的算法决策，人类决策表现出更高的灵活性，故可以接受和理解人类有时做出的不准确决策(Dietvorst & Bharti, 2020)。后续研究发现，算法厌恶广泛存在于金融决策、医疗诊断等领域(Longoni et al., 2019)，且人们拒绝使用算法代替人类进行道德决策(Bigman & Gray, 2018)。Morewedge(2022)进一步提出若算法应用情境的判定标准越模糊(ambiguity of evaluative criteria)且与人类身份相关性(identity relevance)越强，算法厌恶越强烈。因此，算法厌恶的存在意味着人们未必倾向于遵从算法所监控的交通违规行为。

根据上述论述，针对以算法为核心的智能监控，尽管其存在诸多优势，但人们对其既可能欣赏也可能厌恶，故算法监控对交规遵从的影响可能并非总是提升，而是存在边界性，视具体违规情境而定。近年来，研究者试图从解释水平理论(construal level theory)出发，提出人们对行动主体(人类/算法)的解释水平存在不同，且对不同主体实施行为也存在不同的解释水平，当二者匹配时相应的主体则可对作用对象产生最大影响(Trope & Liberman, 2010)。其中，解释水平是指个体对事物表征的抽象程度，即人们对事件或客体的心理表征具有层次性，可分为两个水平：具体而背景化的低水平表征，以及抽象而去背景化的高水平表征(Liberman et al., 2007; Trope & Liberman, 2003, 2010)。例如，一个行为(如粉刷房间)既可被表征为具体的动作行为(如使用刷子涂抹)，即低解释水平，也可被表征为抽象和概要的目标(如让房间变得漂亮)，即高解释水平(Trope & Liberman, 2010)。就行动主体的解释水平而言，Ward 等(2013)发现，由于算法所执行的行为是通过编程来完成的，而不由他们自身的目标或意图驱动，故人们将算法表征为低解释水平的主体，而将具有意图的人类表征为高解释水平的主体。Kim 和 Duhachek(2020)的研究揭示，即使针对相同的行为，被试倾向于使用高解释水平来描述人类行动者，而选择低解释水平来描述算法行动者。因此，人们倾向于将算法表征为重复且不灵活的行动主体，即低解释水平主体，而将人类表征为灵活且具有意图性的高解释水平主体(Loughnan & Haslam, 2007; Kim & Duhachek, 2020)。进一步研究发现，当行动主体执行与之解释水平相匹配的任务时，人们会形成更高的匹配感(即“感觉合适”的体验)，从而受到更大影响(Cesario et al., 2004)。如：相比于高创造性产品(衣服款式)，算法在推荐低创造性产品(衣服尺码)时，人们会感觉到更高的匹配感，从而更倾向于接受算法推荐的衣服尺码而不是衣

服款式(吴继飞 等, 2020)。Kim 和 Duhachek(2020)的研究也发现，对同一件产品，当人类对其进行高解释水平描述而算法进行低解释水平描述时，人们感觉到更匹配，进而更容易被说服购买。

鉴于对行动主体和行为均具有不同的解释水平以及匹配感的作用，本研究认为监控主体(算法监控/人类监控)对交规遵从意愿的影响因交通行为(违规行为)的解释水平而有所差异，且匹配感在其中起中介作用。具体而言，当面对高解释水平的交通行为时，对于同属高解释水平的人类主体，因具有更高的匹配感，从而在该主体监控时人们对交规的遵从高于算法监控；但对低解释水平的交通行为而言，因同属低解释水平的算法主体具有更高的匹配感，从而在该主体监控时人们对交规的遵从高于人类监控。

然而，无论是人类监控还是算法监控都难以完全准确地监控到交通违规行为，常存在监控失误。例如，2020 年郑州市某区因为人类监控或算法监控所产生的失误，驾驶员通过交管 12123 平台提出违法异议的申请超过 6000 起(张笑语, 2021)。那么，当了解监控主体失误后，算法监控和人类监控对不同解释水平交通行为的规则遵从存在怎样的影响呢？本研究将进一步探讨失误后算法监控和人类监控影响交规遵从的差异。研究发现，决策准确性是人类欣赏算法的关键因素，且人们会以更严厉的态度看待算法做出的错误决定(Bogert et al., 2021)。而当算法与人类出现同样的失误时，人们会更快地对算法失去信任(Dietvorst et al., 2015)，且算法监控是否失误对人类的认知和行为影响更大(Mahmud et al., 2022)。据此，本研究提出当人们了解到监控主体的失误信息后，相比人类监控，算法监控效果(即交规遵从)将受到更大影响。其中，低解释水平行为的匹配感被削弱，进而导致算法监控对交规遵从的提升作用变小甚至消失。

为探讨上述所提出的研究问题并检验假设，本研究通过两项研究考察算法监控对交规遵从意愿的影响及其潜在机制。鉴于研究需要编制具体的交通行为情境材料和区分这些行为的解释水平，开展预研究调查常见交通违规行为的解释水平，并筛选出合适的交通违规行为作为后续研究材料。基于预研究所筛选出的交通违规行为，研究 1a 采用情境测验问卷，探讨算法监控对不同解释水平交通行为的规则遵从意愿的影响以及匹配感在其中的作用。为检验研究结论的生态效度，研究 1b 还采用田野调查的方法，根据情境测试材料中描述的交通情境，在现实情境中的相应路口里对交通行为进行观察和记录，旨在检验不同监控主体在现实交通情境中的效果差异。研究二则在研究情境测试材料的基础上，加入监控主体未监测到交通违规行为的失误信息描述，探究存在监控失误时，算法监控对不同解释水平交通行为的规则遵从意愿的影响及其中介机制。

2 预研究：不同解释水平交通违规行为的筛选

预研究旨在选取具有代表性的交通违规行为，并对其解释水平进行评定，为后续研究提供情境材料。

2.1 方法

2.1.1 被试

通过线上广告招募的方式，邀请持有驾照的被试参与纸质问卷的填写，共招募被试 141 名。剔除未填写完整的问卷 9 份，有效问卷 132 份。被试的年龄在 18~50 岁之间($M = 24.05$, $SD = 5.87$)，驾龄在 0.5~28 年之间($M = 2.93$, $SD = 3.20$)。

2.1.2 实验材料与程序

首先，对常见的交通违规行为进行筛选。参考中华人民共和国公安部公开数据，2020 年上半年肇事突出的 10 大交通违规行为分别为：未按规定让行、超速行驶、无证驾驶、醉酒驾驶、未与前车保持安全距离、逆向行驶、违反交通信号、酒后驾驶、违规超车、违规会车(中华人民共和国公安部, 2020)。由于醉酒驾驶与酒后驾驶需通过酒精检测仪测量后判定，在研究中通过行为描述不易操作，因此将醉酒驾驶与酒后驾驶分别替换为驾驶违规行为中常见的未系安全带和用手接打电话；且无证驾驶在行驶过程中也难以判断，故无证驾驶替换为不按规定车道行驶(佟嘉旺, 2013)。

其次，对筛选出的交通违规行为进行解释水平以及普遍性感知评分。对交通违规行为解释水平的评定参考 Vallacher 和 Wegner(1989)编制的行为识别量表(Behavioral Identification Form, BIF)，将调查项目设定为：“你认为以下两种描述，哪种更符合[...根据不同行为调整表述，如：未按规定让行]这一行为？”，回答选项采用-3 至 3 的 7 点评分。其中，“-3”为低解释水平的描述，其指向交通违规行为的具体动作，“3”为高解释水平的描述，其指向交通违规行为的意图原因。例如，对“未按规定让行”，高解释水平的描述为“认为不会撞到行人”，低解释水平的描述为“不减速通过人行横道”。为筛选较为普遍的交通违规行为，参考冯俊辉(2010)编制的《行人交通违规行为调查表》，将项目设置为“根据你的经验，你认为该行为发生的普遍性如何？”项目采用 10 点评分(1 = 几乎没见过，10 = 总是见到)。

在征得被试同意后，使用自编的《交通违规行为调查问卷》进行调查。首先填写被试个人基本信息，之后依次完成对交通违规行为解释水平评定以及不同交通违规行为的普遍性感知评分。

2.2 实验结果与分析

如表 1 所示，高解释水平的交通违规行为依次为：未系安全带、用手接打电话和未按规定让行；低解释水平的交通违规行为依次为：超速行驶、逆向行驶和违规使用灯光。高解释水平的交通违规行为中，用手接打电话和未按规定让行的普遍性感知均较高，而低解释水平的交通违规行为中，超速行驶和违规使用灯光的普遍性感知较高。针对这四种交通行为，分析发现，人们对高解释水平的交通违规行为($M = 6.70$, $SE = 0.14$)与低解释水平的交通违规行为($M = 6.57$, $SE = 0.16$)的普遍性感知无显著差异，$t(131) = 0.80$, $p = 0.428$, Cohen's $d = 0.07$。因此，选择以上四种交通行为作为后续研究的情境材料。

表 1 交通违规行为的解释水平和普遍性评分($M±SE$)

交通违规行为	解释水平评分	普遍性评分
未系安全带	-0.49 ± 0.21	5.04 ± 0.22
用手接打电话	-0.57 ± 0.20	7.05 ± 0.16
未按规定让行	-0.73 ± 0.19	6.35 ± 0.20
未与前车保持安全距	-0.88 ± 0.19	6.19 ± 0.18
不按规定车道行驶	-1.23 ± 0.17	5.19 ± 0.20
违反交通信号	-1.28 ± 0.18	5.10 ± 0.21
违规超车	-1.28 ± 0.18	6.02 ± 0.18
违规使用灯光	-1.42 ± 0.17	6.44 ± 0.20
逆向行驶	-1.44 ± 0.18	4.58 ± 0.22
超速行驶	-1.55 ± 0.17	6.70 ± 0.19

3 研究 1a：算法监控对交规遵从意愿的影响及中介机制

基于预研究筛选的情境材料，研究 1a 旨在探讨相比人类监控，算法监控是否以及如何影响人们对不同解释水平交通行为的规则遵从意愿。

3.1 方法

3.1.1 被试

参考 Bonezzi 等(2022)以及吴继飞等(2020)研究的被试量，计划每个条件至少招募被试 100 名，共计至少 200 名被试。通过线上广告招募的方式，邀请具有驾照的被试参与纸质问卷的填写，共招募被试 217 名。剔除未填写完整的问卷 7 份，最终获得有效被试 210 名。其中，人类监控条件下 107 名(男性 46 名，女性 61 名)，算法监控条件下 103 名(男性 46 名，女性 57 名)。被试年龄在 18~42 岁之间($M = 23.08$，$SD = 3.69$)，驾龄在 0~13 年之间($M = 2.45$, $SD = 2.34$)。

3.1.2 实验设计与程序

实验采用 2(监控主体：人类/算法；被试间) × 2(交通行为解释水平：低解释水平/高解释

水平；被试内)的混合设计。在征得同意后，被试被随机分配到人类监控或算法监控条件下。其中，每种监控主体条件下包含预研究筛选的 2 种解释水平下的 2 个交通行为情境。

高解释水平交通行为(未按规定让行)：试想，你是某一公司员工。今天，你在开车去上班的路上。由于前段道路比较拥堵，耽误了不少时间，导致你上班快要迟到了。你好不容易开出了拥堵路段，十分着急地开车赶着时间。这时，你发现前方路口有行人站在人行横道边准备过马路，且路旁有交警正在执勤(算法监控条件下表述为：且路旁装有电子警察，下同)。

高解释水平交通行为(用手接打电话)：试想，你是某一公司员工。今天，你与一位住在郊区的客户约好了在晚上见面。这位客户是即将与你达成长期合作的重要客户。在开车去客户家的路上，手机铃声突然响起(你没有携带蓝牙耳机且未连接车载蓝牙)。这时，你发现是这位重要客户给你打来的电话，且路旁有交警正在执勤。

低解释水平交通行为(超速行驶)：试想，你是某一公司员工。今天，你在开车去上班的路上。由于前段道路比较拥挤，耽误了不少时间。你好不容易开出了拥堵路段，以约 60km/h 的车速十分着急地开车赶着时间。这时，你发现车速路旁有限速 40km/h 的标志，且路旁有交警正在执勤。

低解释水平交通行为(违规使用灯光)：试想，你是某一公司员工，今天，你与一位住在郊区的客户约好了在晚上见面。这位客户是即将与你达成长期合作的重要客户。在开车去客户家的路上，由于郊区路灯昏暗且弯道较多，你打开了远光灯，以便能看清远方的路况。这时，你发现对面有车辆同样使用远光灯迎面驶来，且路旁有交警正在执勤。

由于被试的驾驶经验、交通法规的熟悉度可能会对交规的理解存在不同，从而影响交规遵从行为，因此在阅读上述材料前，被试需填写驾龄、驾驶里程、驾驶经验与交规熟悉程度等信息。调查项目分别为："驾龄__年，驾驶里程__公里"；"你认为你的驾驶经验如何？请从 0(几乎没有经验)到 100(非常有经验)进行打分___"；"你对交通法规熟悉程度如何？请从 0(几乎不熟悉)到 100(非常熟悉)进行打分___"。

被试在阅读上述材料后，依次完成交规遵从意愿、匹配感和监控准确率感知项目的填写。关于交规遵从意愿的项目，参考 Kim 和 Duhachek (2020)的研究，将其设置为："你估计，在上述情境中，你有多大可能会[…根据所阅读的交通行为情境调整表述，如：等行人过马路]？"项目采用 7 点评分(1 = 几乎不可能，7 = 完全可能)。关于匹配感调查，参考 Kim & Duhachek (2020)和吴继飞等(2020)的研究，包括 3 个项目："你觉得，上述情境中电子警察这一算法监控/交警这一人类监控，对[…根据所阅读的交通行为情境调整表述，如：未按规定让行]这一行为进行监控是合适的/有效的/可信的"。项目采用 7 点评分(1 = 非常不同意，7 = 非常同

意)。就监控准确率感知项目的设置，参考 Kim 和 Duhachek (2020)针对算法/人类推荐专业度的测量，表述为：“你认为，上述情境中电子警察这一算法监控/交警这一人类监控在多大程度上，能准确监测到[...所阅读的交通行为情境调整表述，如：“未按规定让行”]行为？”项目采用 10 点评分(0 = 一点也不准确(0%)，10 = 非常准确(100%))。

3.2 数据整理

测量匹配感 3 个项目一致性较好($0.81 \le \alpha \le 0.90$)，故将匹配感中的项目进行合并分析。高解释水平交通行为下，交规遵从意愿、匹配感和监控准确率感知相关系数别为 0.19 (p = 0.007)、0.46 ($p < 0.001$)和 0.35 ($p < 0.001$)；低解释水平交通行为下，交规遵从意愿、匹配感和监控准确率感知相关系数分别为 0.34 ($p < 0.001$)、0.47 ($p < 0.001$)和 0.41 ($p < 0.001$)。据此，将不同解释水平交通行为下两个情境中对应项目的平均分作为后续统计分析的数据。

3.3 结果与分析

不同监控主体和交通行为解释水平下，被试的交规遵从意愿评分如图 1a 所示。2(监控主体：人类/算法；被试间) × 2(交通行为解释水平：低解释水平/高解释水平；被试内)混合设计的方差分析结果发现，监控主体的主效应不显著，$F(1, 208) = 0.03$, $p = 0.875$, $\eta_p^2 < 0.01$；交通行为解释水平的主效应显著，$F(1, 208) = 76.75$, $p < 0.001$, $\eta_p^2 = 0.27$，即被试在低解释水平条件下的交规遵从意愿($M = 6.16$, $SE = 0.07$)显著高于高解释水平条件($M = 5.25$, $SE = 0.08$)；两者交互作用显著，$F(1, 208) = 19.78$, $p < 0.001$, $\eta_p^2 = 0.09$。简单效应分析结果显示，在高解释水平交通行为条件下，人类监控时被试的交规遵从意愿($M = 5.49$, $SE = 0.12$)显著高于算法监控($M = 5.01$, $SE = 0.12$；$t(208) = 2.89$, $p = 0.004$, Cohen’s $d = 0.32$)；在低解释水平交通行为条件下，人类监控时被试的交规遵从意愿($M = 5.94$, $SE = 0.10$)显著低于算法监控条件($M = 6.38$, $SE = 0.11$；$t(208) = -3.00$, $p = 0.003$, Cohen’s $d = -0.30$)。因此，监控主体影响人们对不同解释水平交通行为的交规遵从意愿；对于低解释水平交通行为而言，算法监控可提高人们的交规遵从意愿，而对于高解释水平交通行为而言，人类监控可提高人们的交规遵从意愿。

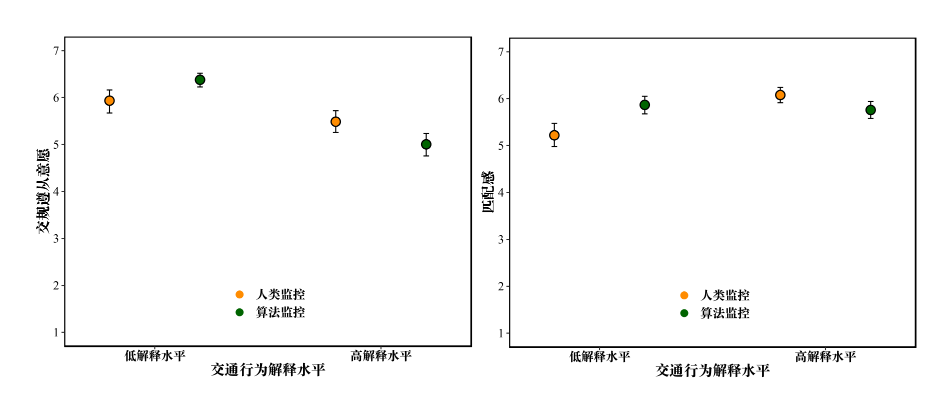


图 1 研究 1a 中不同条件下的交规遵从意愿与匹配感 ($M \pm SE$)

不同监控主体和交通行为解释水平下，被试所报告的匹配感评分如图 1b 所示。2(监控主体：人类/算法；被试间) × 2(交通行为解释水平：低解释水平/高解释水平；被试内)混合设计的方差分析结果发现，监控主体的主效应不显著，$F(1, 208) = 1.70, p = 0.194, \eta_p^2 < 0.01$；交通行为解释水平的主效应显著，$F(1, 208) = 23.64, p < 0.001, \eta_p^2 = 0.10$，即被试在高解释水平下的匹配感($M = 5.92, SE = 0.07$)显著高于低解释水平($M = 5.54, SE = 0.08$)；两者交互作用显著，$F(1, 208) = 39.29, p < 0.001, \eta_p^2 = 0.16$。简单效应分析结果显示，在高解释水平交通行为条件下，人类监控下被试的匹配感($M = 6.08, SE = 0.09$)显著高于算法监控($M = 5.76, SE = 0.09$；$t(208) = 2.47, p = 0.014$, Cohen's $d = 0.29$)；但在低解释水平交通行为条件下，人类监控时被试的匹配感($M = 5.22, SE = 0.12$)显著低于算法监控条件($M = 5.87, SE = 0.12$；$t(208) = 3.94, p < 0.001$, Cohen's $d = 0.58$)。该结果模式与交规遵从意愿结果相似。

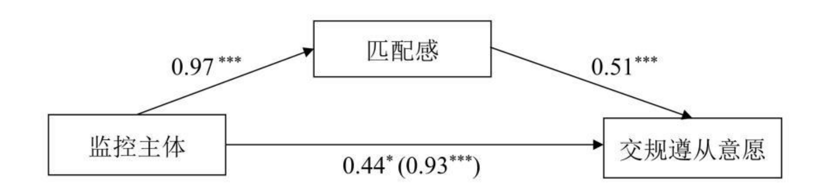


图 2 研究 1a 中匹配感在监控主体与交规遵从意愿之间的中介效应图

为进一步检验匹配感在交规遵从意愿中的作用，鉴于交通行为的解释水平是被试内变量，参考以往研究(Bonezzi et al., 2022)，分别将低解释水平交通行为下的匹配感和交规遵守意愿评分减去两者在高解释水平交通行为下的评分，将得到的匹配感差值作为中介变量，交规遵从意愿差值作为因变量(为表述简洁化，以下依然使用匹配感和交规遵从意愿作为变量术语)，以监控主体为自变量(人类监控 = 0，算法监控 = 1)进行中介效应分析。结果发现，该模型的中介效应值为 $0.97 \times 0.51 = 0.49$。采用偏差校正的 Bootstrap 方法对中介效应进行显著性检验(抽取 5000 次)，发现 95%的置信区间为[0.24, 0.85]，不包含 0，表明中介作用显著，其中介效应路径图见图 2。这一分析结果表明，监控主体可通过影响匹配感从而影响道路交通情境中的交规遵守意愿。

表 2 研究 1a 中不同监控主体和解释水平下监控准确率感知

交通行为解释水平	监控主体	*M*	*SE*
低解释水平	人类监控	6.30	0.20
	算法监控	7.71	0.16
高解释水平	人类监控	6.85	0.16
	算法监控	7.46	0.16

针对上述所有分析，将驾龄、驾驶里程、驾驶经验、交规熟悉程度作为协变量控制后，实验结果依然相同，且不能由监控准确率感知差异所解释。具体而言，对监控准确率感知进行2(监控主体：人类/算法；被试间)×2(交通行为解释水平：低解释水平/高解释水平；被试内)混合设计的方差分析，虽然发现交互作用显著，$F(1, 208) = 5.88, p = 0.015, \eta_p^2 = 0.03$，但简单效应分析发现，无论是低解释水平($t(208) = 5.39, p < 0.001$, Cohen's $d = 0.60$)还是高解释水平($t(208) = 2.68, p = 0.008$, Cohen's $d = 0.26$)的交通行为条件下，被试对算法监控准确率的感知均高于人类监控(描述性统计见表2)。

4 研究1b：算法监控影响交规遵从的田野调查

为检验研究1a结论的生态效度，研究1b采用田野调查的方法，根据研究1a中情境测试材料所描述的交通情境，在现实情境中找到相应的路口，并对交通行为进行观察和记录。

4.1 设计与程序

基于研究1a文字材料所描述的情境，本研究拟在以下四种现实交通情境中观察不同交通违规行为出现的次数，其中算法监控路口的选取以“高德地图”导航提示为依据。具体情境包括：没有任何监控的路口、只有交警监控的路口、只有算法监控的路口、既有交警也有算法监控的路口。

选取周四上班早高峰时段内2小时，即上午7:30至9:30作为观察时段，共观察两次。每观察2分钟记为观察1轮，对每个交通情境路口在选取的时间段内共观察10轮，即共观察20分钟。观察内容以手机视频录制的形式进行记录，之后由两名编码者基于录制的视频编码不同解释水平交通违规行为出现的次数。同时，鉴于拥挤及混乱的交通流会影响驾驶员的心理状态，使其容易产生烦躁、易怒等负面情绪，进而致使驾驶员做出交通违规行为，因此本研究将道路交通流量(即每分钟内通过的车辆数)纳入观察指标进行分析。

4.2 数据整理与统计预分析

针对每个情境中出现的交通违规行为，由两名持有驾照的编码者依据交通法规作为编码标准，对其出现次数进行归类编码。两名编码者对交通违规行为次数的编码有较高一致性(CA = 0.88)，且编码可靠(R = 0.98)。对于少部分编码存在分歧的交通违规行为，由两名编码者讨论后共同确定其归类。编码结果发现，所有路口均出现次数不等的未按规定让行这一高解释水平的交通违规行为，但未观察到其他交通违规行为。

4.3 结果与分析

由于每组(即条件)的观察样例较少($n = 10$)，因此采用Kruskal-Wallis H非参数检验比较

不同路口类型下的道路交通流量与交通违规行为次数。结果显示，不同路口类型下的道路交通流量差异显著，$H = 26.62$, $p < 0.001$。事后多重比较结果(FDR 校正)表明，在只有交警监控的路口(*Median, MD* = 23.63)、只有算法监控的路口(*MD* = 17.75)、既有交警也有算法监控的路口(*MD* = 22.00)之间道路交通流量无显著差异($p_s > 0.053$)，但均高于没有任何监控的路口(*MD* = 13.00)的道路交通流量($p_s < 0.024$)。该结果说明，存在监控的三种路口之间道路交通流量基本相当，故后续发现交通违规行为的差异并不能由道路交通流量的差异所解释。但由于在现实情境中，车流量较少的路段常常不设置监控，因此没有任何监控的路口的道路交通流量低于其他情境的路口，该差异符合现实情况。

不同路口类型下的高解释水平交通违规行为的出现次数差异显著，$H = 21.26$, $p < 0.001$。事后多重比较结果(FDR 校正)表明，在只有交警监控的路口(*Median, MD* = 0.00)和既有交警也有算法监控的路口(*MD* = 2.00)，高解释水平交通违规行为的出现次数显著低于没有任何监控的路口(*MD* = 5.00; $p < 0.001$, $p = 0.044$)，而在只有算法监控的路口(*MD* = 5.00)，高解释水平交通违规行为的出现次数与没有任何监控的路口(*MD* = 5.00)之间无显著差异($p = 1.000$)。进一步比较发现，在只有交警监控的路口和既有交警也有算法监控的路口，高解释水平交通违规行为的出现次数显著低于只有算法监控的路口($p < 0.001$; $p = 0.044$)。以上结果验证了研究 1a 的部分结论，对于高解释水平交通行为，相比于算法监控，人类监控更能够提升人们的交规遵从(违规行为出现次数见图 3)。

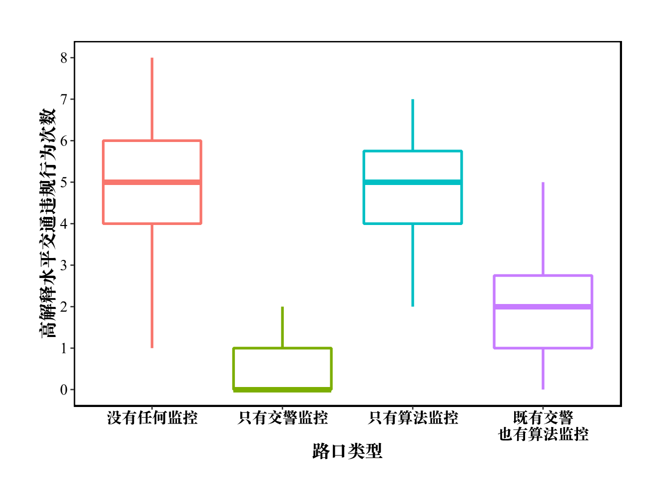


图 3 不同路口类型下高解释水平交通违规行为的出现次数

5 研究 2a：存在失误的算法监控对交规遵从意愿的影响

研究 2a 旨在探讨当监控存在失误时，相比人类监控，算法监控对不同解释水平交通行为的交规遵从意愿的影响差异。

5.1 方法

5.1.1 被试

参考 Bonezzi 等(2022)以及吴继飞等(2020)研究的被试数量，计划至少招募被试 200 名，其中每个条件下至少 100 名。通过线上广告招募的方式，邀请持有驾照的被试参与纸质问卷的填写。共招募被试 214 名，均填写完整。其中，人类监控条件下 106 名(男性 54 名，女性 52 名)，算法监控条件下 108 名(男性 55 名，女性 53 名)。被试年龄在 19~24 岁之间(M = 21.43，SD = 1.65)，驾龄在 1~4 年之间(M = 2.20, SD = 1.09)。

5.1.2 实验设计与程序

实验采用 2(监控主体：人类/算法；被试间) × 2(交通行为解释水平：低解释水平/高解释水平；被试内)的混合设计。

研究 2a 的流程与研究 1a 相似，但交通行为情境材料进行了调整。具体而言，交通行为情境材料中增加监控主体失误的描述，即过去有驾驶员未遵守交通规则，但并未被监控主体监测到。例如，未按规定让行情境中，对人类监控而言，在研究 1a 交通行为情境材料的最后增加“据了解，虽然有驾驶员未按规定给行人让行，但仍有少数未被路旁的这名交警监测到。”；对算法监控而言，增加“据了解，虽然有驾驶员未按规定给行人让行，但仍有少数未被路旁的这个电子警察监测到。”

被试阅读完材料后，依次完成交规遵从意愿和监控准确率感知的题项。调查项目的设置同研究 1a。

5.2 数据整理

高解释水平交通行为下，交规遵从意愿和监控准确率感知相关系数分别为 0.28 (p < 0.001)和 0.56 (p < 0.001)；低解释水平交通行为下，交规遵从意愿和监控准确率感知相关系数分别为 0.61 (p < 0.001)和 0.58 (p < 0.001)。据此，将不同解释水平交通行为下两个情境中对应项目的平均分作为后续统计分析的数据。

5.3 结果与分析

不同监控主体和交通行为解释水平下的交规遵从意愿评分如图 4 所示。2(监控主体：人类/算法；被试间) × 2(交通行为解释水平：低解释水平/高解释水平；被试内)混合设计的方差分析结果发现，监控主体的主效应显著，$F(1, 212) = 3.94$, $p = 0.048$, $\eta_p^2 = 0.02$，即被试在人类监控下的交规遵从意愿(M = 4.62, SE = 0.11)显著高于算法监控(M = 4.28, SE = 0.10)；交通行为解释水平的主效应不显著，$F(1, 212) = 0.42$, $p = 0.52$, $\eta_p^2 = 0.002$；两者交互效应显著，$F(1, 212) = 4.01$, $p = 0.046$, $\eta_p^2 = 0.02$。简单效应分析结果显示，在高解释水平交通行为条件

下，人类监控时被试的交规遵从意愿(M = 4.70, SE = 0.14)显著高于算法监控(M = 4.12, SE = 0.14；t(212) = 2.98, p = 0.003, Cohen's d = 0.40)；在低解释水平交通行为条件下，人类监控时被试的交规遵从意愿(M = 4.53, SE = 0.16)与算法监控条件(M = 4.44, SE = 0.15)之间无显著差异，t(212) = 0.40, p = 0.688, Cohen's d = 0.06。因此，监控主体的失误信息影响人们对不同解释水平交通行为的遵守意愿；对高解释水平交通行为而言，人类监控依然可提高交规遵从意愿，但对于低解释水平交通行为而言，算法监控与人类监控对交规遵从意愿的影响相当。

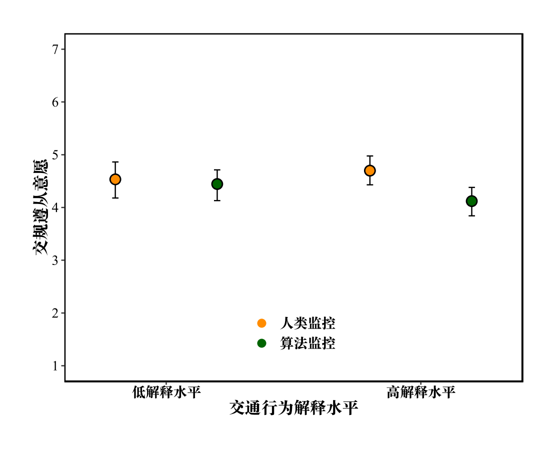


图 4 研究 2a 中不同条件下的交规遵从意愿(M±SE)

针对上述分析，将驾龄、驾驶经验、交规熟悉程度作为协变量控制后，实验结果依然相同，且不能由监控准确率感知差异所解释。具体而言，对监控准确率感知进行 2(监控主体：人类/算法；被试间) × 2(交通行为解释水平：低解释水平/高解释水平；被试内)混合设计的方差分析，虽然发现交互作用显著，F(1, 212) = 4.01, p = 0.046, η_p^2 = 0.02，但简单效应分析发现，无论是低解释水平(t(212) = 1.23, p = 0.220, Cohen's d = 0.13)还是高解释水平(t(212) = 0.17, p = 0.866, Cohen's d = 0.02)的交通行为条件下，被试对不同监控主体型的监控准确率的感知均无显著差异(见表 3)。

表 3 研究 2 中不同监控主体和解释水平下监控准确率感知

研究	交通行为解释水平	监控主体	M	SE
研究 2a	低解释水平	人类监控	5.73	0.23
		算法监控	6.13	0.23
	高解释水平	人类监控	5.96	0.23
		算法监控	6.01	0.21
研究 2b	低解释水平	人类监控	7.00	0.16
		算法监控	7.49	0.16
	高解释水平	人类监控	6.34	0.14
		算法监控	7.20	0.14

5.4 跨研究比较

为进一步探讨研究 1a 与研究 2a 的交规遵从意愿模式是否存在不同影响，对结果进行了跨研究比较。以交规遵从意愿作为因变量，分别对不同解释水平交通行为进行 2(交通情境类型：无失误信息(研究 1a)/有失误信息(研究 2a)) × 2(监控主体：人类/算法)完全被试间方差分析。

对于高解释水平的交通行为，交通情境类型的主效应显著，$F(1, 420) = 42.67, p < 0.001, \eta_p^2 = 0.09$，即被试在了解监控主体有失误经历后交规遵从意愿($M = 4.41, SE = 0.10$)显著低于监控主体无失误描述的情况($M = 5.25, SE = 0.85$)；监控主体的主效应显著，$F(1, 420) = 17.11, p < 0.001, \eta_p^2 = 0.04$，即在人类监控条件下，被试的交规遵从意愿($M = 5.09, SE = 0.94$)显著高于算法监控条件($M = 4.55, SE = 0.96$)；两者交互效应不显著，$F(1, 420) = 0.14, p = 0.706, \eta_p^2 < 0.01$。

对于低解释水平的交通行为，交通情境类型的主效应显著，$F(1, 420) = 157.10, p < 0.001, \eta_p^2 = 0.27$，即被试在监控主体有失误经历后交规遵从意愿($M = 4.49, SE = 0.11$)显著低于对监控主体无失误描述的情况($M = 6.15, SE = 0.75$)；监控主体的主效应不显著，$F(1, 420) = 1.78, p = 0.182, \eta_p^2 < 0.01$；两者交互效应显著，$F(1, 420) = 4.01, p = 0.046, \eta_p^2 = 0.01$。简单效应分析结果显示，在低解释水平的交通行为下，当未提供监控主体失误信息时，算法监控条件下被试的交规遵从意愿($M = 6.38, SE = 0.08$)显著高于人类监控条件($M = 5.93, SE = 0.13$; $t(208) = 3.00, p = 0.003$, Cohen's $d = 0.42$)；但当提供监控主体存在失误经历后，算法监控条件下被试的交规遵从意愿($M = 4.44, SE = 0.15$)与人类监控条件($M = 4.53, SE = 0.16$)之间无显著差异，$t(212) = 0.40, p = 0.688$, Cohen's $d = 0.06$。

本研究表明，当了解监控主体的失误信息后，对高解释水平交通行为而言，相比算法监控，人类监控依然可提高人们的交规遵从意愿，但对于低解释水平交通行为而言，与研究 1a 不同，相比人类监控，算法监控对交规遵从意愿的提升作用消失。

6 研究 2b：存在失误的算法监控对交规遵从意愿的影响及中介机制

研究 2b 旨对研究 2a 进行重复，并检验匹配感是否在算法监控影响不同解释水平交通行为的交规遵从意愿中起中介作用。

6.1 方法

被试量的确定以及招募与研究 2a 相同，共招募被试 223 名。剔除未填写完整的问卷 9 份，有效问卷 214 份。其中，人类监控条件下 107 名(男性 68 名，女性 39 名)，算法监控条

件下 107 名(男性 76 名，女性 31 名)。被试年龄在 18~35 岁之间(M = 22.11，SD = 2.67)，驾龄在 0~11 年之间(M = 2.26, SD = 1.70)。

研究 2b 的流程与研究 2a 相似，但调查项目的设置加入了匹配感的测量，其同研究 1a。

6.2 结果与分析

不同监控主体和交通行为解释水平下，被试的交规遵从意愿评分如图 5 所示。2(监控主体：人类/算法；被试间) × 2(交通行为解释水平：低解释水平/高解释水平；被试内)混合设计的方差分析结果发现，监控主体的主效应显著，$F(1, 212) = 11.58, p < 0.001, \eta_p^2 = 0.05$，即被试在人类监控下的交规遵从意愿($M$ = 4.99, SE = 0.09)显著高于算法监控(M = 4.58, SE = 0.09)；交通行为解释水平的主效应显著，$F(1, 212) = 4.78, p = 0.030, \eta_p^2 = 0.02$，显示被试对低解释水平交通行为的交规遵从意愿($M$ = 4.92, SE = 0.10)显著高于高解释水平(M = 4.66, SE = 0.07)；两者交互效应显著，$F(1, 212) = 8.62, p = 0.004, \eta_p^2 = 0.04$。简单效应分析结果显示，在低解释水平交通行为条件下，人类监控时被试的交规遵从意愿(M = 4.95, SE = 0.14)与算法监控条件(M = 4.89, SE = 0.14)之间无显著差异，$t(212) = 0.32, p = 0.751$, Cohen's d = 0.04；在高解释水平交通行为条件下，人类监控时被试的交规遵从意愿(M = 5.04, SE = 0.10)显著高于算法监控(M = 4.28, SE = 0.10；$t(212) = 5.31, p < 0.001$, Cohen's d = 0.44)。因此，对于高解释水平交通行为而言，人类监控依然可提高人们的交规遵从意愿，但对于低解释水平交通行为而言，算法监控与人类监控对交规遵从意愿的影响相当，重复了研究 2a 的结果。

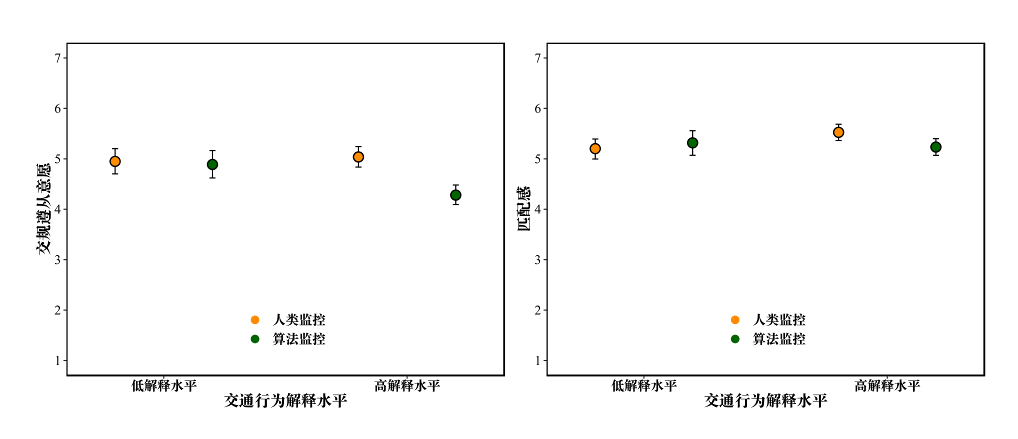


图 5 研究 2b 中不同条件下的交规遵从意愿与匹配感(M±SE)

不同监控主体和交通行为解释水平下，被试所报告的匹配感评分如图 5 所示。方差分析结果发现，监控主体的主效应不显著，$F(1, 212) = 0.78, p = 0.379, \eta_p^2 < 0.01$，交通行为解释水平的主效应不显著，$F(1, 212) = 1.50, p = 0.223, \eta_p^2 < 0.01$，但两者交互作用显著，$F(1, 212) = 4.26, p = 0.040, \eta_p^2 = 0.02$。简单效应分析结果显示，在低解释水平交通行为条件下，人类监控的匹配感(M = 5.20, SE = 0.11)与算法监控(M = 5.32, SE = 0.11；$t(212) = -0.73, p = 0.468$,

Cohen's $d = 0.08$)差异不显著；但在高解释水平交通行为条件下，人类监控的匹配感($M = 5.53$, $SE = 0.09$)显著高于算法监控($M = 5.23$, $SE = 0.09$；$t(212) = 2.43$, $p = 0.016$, Cohen's $d = 0.20$)。该结果模式与研究 2a 交规遵从意愿相似。

采用研究 1a 相同的中介效应分析方法发现，模型的中介效应值为 0.41 × 0.48 = 0.19。采用偏差校正的 Bootstrap 方法对中介效应进行显著性检验(抽取 5000 次)，发现 95%的置信区间为[0.02, 0.40]，不包含 0，表明中介作用显著，其中介效应路径图见图 6。这一分析结果表明，当监控失误后，监控主体依然可通过影响匹配感从而影响道路交通情境中的交规遵守意愿。

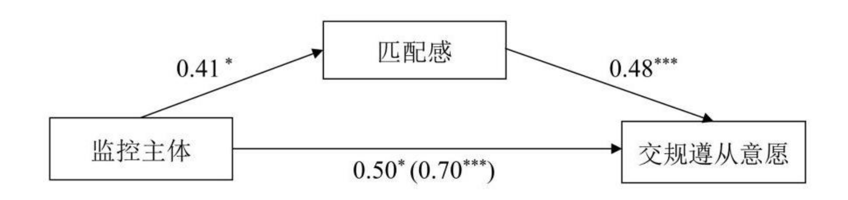


图 6 研究 2b 中匹配感在监控主体与交规遵从意愿之间的中介效应图

针对上述所有分析，将驾龄、驾驶经验、交规熟悉程度作为协变量控制后，实验结果依然相同，且不能由监控准确率感知差异所解释。虽然针对监控准确率感知的 2(监控主体：人类/算法；被试间) × 2(交通行为解释水平：低解释水平/高解释水平；被试内)方差分析的交互作用并不显著，$F(1, 212) = 2.78$, $p = 0.097$, $\eta_p^2 = 0.01$，但无论是低解释水平($t(212) = 2.16$, $p = 0.032$, Cohen's $d = 0.29$)还是高解释水平($t(212) = 4.25$, $p < 0.001$, Cohen's $d = 0.52$)的交通行为条件下，被试对算法监控的准确率感知均高于人类监控(见表 3)。

6.3 跨研究比较

与研究 2a 相同，以交规遵从意愿作为因变量，分别对不同解释水平交通行为进行 2(交通情境类型：无失误信息(研究 1a)/有失误信息(研究 2b)) × 2(监控主体：人类/算法)完全被试间方差分析。

对于高解释水平的交通行为，交通情境类型的主效应显著，$F(1, 420) = 28.73$, $p < 0.001$，即被试在了解监控主体有失误经历后交规遵从意愿($M = 4.66$, $SE = 0.08$)显著低于对监控主体无失误描述的情况($M = 5.25$, $SE = 0.08$)；监控主体的主效应显著，$F(1, 420) = 32.01$, $p < 0.001$, $\eta_p^2 = 0.07$, $\eta_p^2 = 0.06$，即在人类监控条件下，被试的交规遵从意愿($M = 5.26$, $SE = 0.08$)显著高于算法监控条件($M = 4.64$, $SE = 0.08$)；两者交互效应不显著，$F(1, 420) = 1.59$, $p = 0.208$, $\eta_p^2 < 0.01$。

对于低解释水平的交通行为，交通情境类型的主效应显著，$F(1, 420) = 104.31$, $p < 0.001$,

$\eta_p^2 = 0.20$，即监控主体有失误经历后被试的交规遵从意愿($M = 4.92$, $SE = 0.11$)显著低于对监控主体无失误描述的情况($M = 6.16$, $SE = 0.09$)；监控主体的主效应不显著，$F(1, 420) = 2.50$, $p = 0.115$, $\eta_p^2 < 0.01$；两者交互效应显著，$F(1, 420) = 4.33$, $p = 0.038$, $\eta_p^2 = 0.01$。简单效应分析结果显示，在低解释水平的交通行为下，当未提供监控主体失误信息时，人类监控时被试的交规遵从意愿($M = 5.94$, $SE = 0.10$)显著低于算法监控条件($M = 6.38$, $SE = 0.11$；$t(208) = -3.00$, $p = 0.003$, Cohen's $d = -0.30$)；但当了解监控失误后，人类监控下的交规遵从意愿($M = 5.20$, $SE = 0.11$)与算法监控($M = 5.32$, $SE = 0.11$；$t(212) = -0.73$, $p = 0.468$, Cohen's $d = 0.08$)差异不显著。

本研究重复了研究 2a 的结果，且发现存在监控失误时，监控主体依然可通过匹配感进而影响道路交通情境中的交规遵从意愿。

7 讨论

针对算法监控如何影响交规遵守的问题，考虑到人们对算法和人类主体实施行为的解释水平存在差异，本研究提出算法监控与所监控的行为在解释水平上匹配时，其对交规遵守意愿产生的影响更强。结果发现，对低解释水平的交通行为，相比人类监控，人们对算法监控形成的匹配感更强，进而提升了对交规的遵守意愿；而对高解释水平的交通行为，人类监控下人们的匹配感更强，从而有更强的意愿遵守交规，且在真实的交通情境也可观察到类似的结果模式。进一步研究发现，当了解算法失误后，其诱发的匹配感降低至与人类监控相当水平，且对低解释水平交通行为遵守意愿的提升作用消失，但人类监控失误时，其对高解释水平交通行为的监控效果依然优于算法监控，即算法失误削弱了其对低解释水平交通行为的监控效果。

从算法对交通违规行为的监控效果来看，人们对算法既表现出欣赏(即相比人类监控交规遵守意愿提升)，又表现出厌恶(即相比人类监控交规遵守意愿下降)，其视不同监控行为而定。本研究拓展了以往对算法厌恶和欣赏看似矛盾的两类发现，揭示至少在交通行为领域，算法欣赏还是厌恶与主体实施行为的解释水平有关。这与算法说服领域中 Kim 和 Duhachek(2020)发现一致。其中，当智能算法采用低解释水平的说服信息来推荐产品(如：如何使用茶树精油)时，说服效果好于采用高解释水平的说服信息(如：为什么要使用茶树精油)，而当智能算法被描述为具有人类的特性(如：具有一定的自我意识)时，前述结果反转。因此，算法所作用或监控行为的解释水平可能是决定厌恶或欣赏的重要变量。究其原因，与人们对算法和人类主体的解释水平有关。毕竟，人类拥有意识和意图且可自主学习，而算法仅仅是

人类设计的程序，并不像人类那样拥有意识和自主学习的能力(Cadario et al., 2021)。人们对算法适合执行哪些行为具有一定的预期，当其与实际执行的行为吻合时产生更高匹配感，进而发挥更强效用，呈现算法欣赏；反之，则表现出算法厌恶。本研究也验证了匹配感在算法监控影响人类交通行为中的中介作用。已有大量研究证明了匹配感在行为选择中的作用，如Kamins(1990)发现当代言人形象与产品形象相匹配时，消费者会有更高的匹配感，从而对广告内容产生信任，使得代言人对消费者的购买态度和行为产生更积极的影响；Lee 和Aaker(2004)的研究也发现，当信息框架(获益与损失)与人们的调节聚焦(促进与预防)相匹配时，人们感到更高的流畅性，从而形成更高的匹配感，并提升了信息说服的效果。总之，算法对个体行为产生影响的效果取决于该行为的解释水平，且匹配感可中介这一作用过程。

监控失误不仅使交规遵从意愿整体变弱，且针对低解释水平的交通行为，相比人类监控，算法的监控效果发生了变化，即算法监控对交规遵从意愿的提升作用消失。这与以往关于算法失误的相关研究一致，其揭示算法失误通常是系统的且难以修复，人们认为算法难以像人类那样从错误中学习并改进，故对算法失误容忍度更低，进而削弱算法对人类行为的作用效果(Mahmud et al., 2022; Cadario et al., 2021; Dietvorst et al., 2015)。从本研究匹配感的结果来看，算法失误也削弱了算法监控与低解释水平交通行为的匹配感，提示人们对算法监控的已有表征或预期较易受到算法失误的影响，故人们更依赖于算法的实际表现形成对算法的认识。对于人类的行为来说，人们对其认识和表征不仅仅建立在观察到的实际表现，且较多依赖于先验知识(prior knowledge)，如人类行为的经济性和灵活性(Jara-Ettinger et al., 2016; Ionescu et al., 2017)。这种灵活性等先验知识具有一定的稳定性，较难在短时间内改变(Carruthers et al., 2008)，因此即使获知人类失误，人们依然会基于原有的知识来形成人类主体与监控行为的匹配感，从而呈现与未获知任何失误信息时相似的结果模式。

研究发现在高解释水平的交通行为中，虽然人们认为算法监控的准确性高于人类监控，但后者可提升对交规遵从意愿。因此，面对不同类型监控，人们选择交规遵从行为时不只是考虑其功效性或有用性，还可能审视人类特有的某些属性。Morewedge(2022)提出算法的影响程度取决于任务或决策情境的两个维度：判定标准的模糊性和与人类身份的相关性。人类的独特行为在这两个维度上均具有较高的得分，故若两者均高则呈现算法厌恶。交通违规行为从其判定模糊性来说，即存在较为清晰的超速，也存在较为模糊的让行人；从其与人类身份的相关性的角度来看，即存在涉及人类独特性的意图性，也具有机械的物理运动。针对本研究采用的两例高解释水平的交通行为(未按规定让行与用手接打电话)，违规判断标准的模糊性和涉及的意图性识别远超低解释水平的两例交通行为(超速行驶与违规使用灯光)，故这

种偏向人类独特性的交通行为使得被试不只是考虑监控的准确性。Dietvorst 和 Bharti(2020)提出，这种独特性认识的关键是人类决策灵活性致使决策误差的波动性也较大，而这种波动性存在一定的概率让人类的决策质量接近完美，同时人们对接近完美答案的决策具有较强的偏好性。

本研究的结论可帮助指导如何利用当前的人工智能算法促进人们对交规的遵从。首先，算法监控对交通规则遵从意愿存在积极影响，为使用电子警察这类算法监控提供了实证支撑。其次，针对监控主体对不同解释水平交通行为的规则遵从存在不同影响的结果，提示交通部门在选择设置监控主体时，应避免在路口无差别地使用算法监控或人类监控，需要考虑所在路口常见的交通违规行为类型。例如，在常出现不礼让行人的重要路口应考虑安排一定的辅警执勤，在常出现超速行驶的路口可以算法监控为主。最后，鉴于监控主体失误后算法监控效果削弱的发现，建议交通部门在违章监测漏报较多的电子警察监控路口灵活安排辅警执勤，以降低算法失误所带来的消极影响。当然，未来研究可招募驾驶经验更丰富的被试，并选取交通违规程度不同的城市进行田野调查，以进一步检验研究结论的普遍性和稳定性。

8 结论

本研究探讨了算法监控对交规遵从意愿的影响及潜在机制，获得以下结论：算法监控对交规遵从意愿的提升作用视交通行为解释水平而定，且匹配感在其中起中介作用；算法失误削弱其对低解释水平交通行为的监控效果。

参考文献

Bigman, Y. E., & Gray, K. (2018). People are averse to machines making moral decisions. *Cognition, 181*, 21–34.

Bigman, Y. E., Wilson, D., Arnestad, M. N., Waytz, A., & Gray, K. (2023). Algorithmic discrimination causes less moral outrage than human discrimination. *Journal of Experimental Psychology: General, 152*(1), 4–27.

Bigman, Y. E., Yam, K. C., Marciano, D., Reynolds, S. J., & Gray, K. (2021). Threat of racial and economic inequality increases preference for algorithm decision-making. *Computers in Human Behavior, 122*, 106859.

Bogert, E., Schecter, A., & Watson, R. T. (2021). Humans rely more on algorithms than social influence as a task becomes more difficult. *Scientific Reports, 11*(1), 8028.

Bonezzi, A., Ostinelli, M., & Melzner, J. (2022). The human black-box: The illusion of understanding human better than algorithmic decision-making. *Journal of Experimental Psychology: General, 151*(9), 2250–2258.

Cadario, R., Longoni, C., & Morewedge, C. K. (2021). Understanding, explaining, and utilizing medical artificial intelligence. *Nature Human Behaviour, 5*(12), 1636–1642.

Carruthers, P., Laurence, S., & Stich, S. (2008). The Innate Mind: volume 3: foundations and the future. *Oxford University Press.*

Cesario, J., Grant, H., & Higgins, E. T. (2004). Regulatory fit and persuasion: Transfer from “feeling right”. *Journal of Personality and Social Psychology, 86*(3), 388–404.

Clarke, R. (1988). Information technology and dataveillance. *Communications of the ACM, 31*(5), 498–512.

Dietvorst, B. J., & Bharti, S. (2020). People reject algorithms in uncertain decision domains because they have diminishing sensitivity to forecasting error. *Psychological Science, 31*(10), 1302–1314.

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General, 144*(1), 114–126.

Feng, J. H. (2010). Research of pedestrian traffic violations based on psychology. *Kunming University of Science and Technology.*

[冯俊辉. (2010). 基于心理学的行人交通违规行为的研究(硕士学位论文). *昆明理工大学*.]

Genda, E. A., Hakim, A., Noor, I., & Setyowati, E. (2023). Electronic Traffic Law Enforcement in Indonesia. *KnE Social Sciences,* 982-990.

Herath, H. M. K. K. M. B., & Mittal, M. (2022). Adoption of artificial intelligence in smart cities: A comprehensive review. *International Journal of Information Management Data Insights, 2*(1), 100076

Hertz, N., Shaw, T., De Visser, E. J., & Wiese, E. (2019). Mixing it up: How mixed groups of humans and machines modulate conformity. *Journal of Cognitive Engineering and Decision Making, 13*(4), 242–257.

Hoc, J. M. & Lemoine, M. P. (2023). Cognitive evaluation of human–human and human–machine cooperation modes in air traffic control. *International Journal of Aviation Psychology, 8*(1), 1–32.

Ionescu, T. (2017). The variability-stability-flexibility pattern: A possible key to understanding the flexibility of the human mind. *Review of General Psychology, 21*(2), 123–131.

Jara-Ettinger, J., Gweon, H., Schulz, L. E., & Tenenbaum, J. B. (2016). The Naïve Utility Calculus: Computational Principles Underlying Commonsense Psychology. *Trends in Cognitive Sciences, 20*(8), 589–604.

Kamins, & Michael, A. (1990). An investigation into the “match-up” hypothesis in celebrity advertising: when beauty may be only skin deep. *Journal of Advertising, 19*(1), 4–13.

Karimi, F., & Matous, P. (2018). Mapping diversity and inclusion in student societies: A social network perspective. *Computers in Human Behavior, 88*, 184–194.

Kawaguchi, K. (2021). When will workers follow an algorithm? A field experiment with a retail business. *Management Science, 67*(3), 1670–1695.

Kim, T. W., & Duhachek, A. (2020). Artificial intelligence and persuasion: A construal-level account. *Psychological Science, 31*(4), 363–380.

Krügel, S., Ostermaier, A., & Uhl, M. (2023). Algorithms as partners in crime: A lesson in ethics by design. *Computers in Human Behavior, 138*, 107483.

Laï, M. C., Brian, M., & Mamzer, M. F. (2020). Perceptions of artificial intelligence in healthcare: findings from a qualitative survey study among actors in France. *Journal of Translational Medicine, 18*(1), 1–13.

Lee, A. Y., & Aaker, J. L. (2004). Bringing the frame into focus: The influence of regulatory fit on processing fluency and persuasion. *Journal of Personality and Social Psychology, 86*(2), 205–218.

Leib, M., Köbis, N., Soraperra, I., Weisel, O., & Shalvi, S. (2021). Collaborative dishonesty: A meta-analytic review. *Psychological Bulletin, 147*(12), 1241.

Liberman, N., Trope, Y., McCrea, S. M., & Sherman, S. J. (2007). The effect of level of construal on the temporal distance of activity enactment. *Journal of Experimental Social Psychology, 43*(1), 143–149.

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes, 151*, 90–103.

Longoni, C., Bonezzi, A., & Morewedge, C. K. (2019). Resistance to medical artificial intelligence. *Journal of Consumer Research, 46*(4), 629–650.

Loughnan, S., & Haslam, N. (2007). Animals and androids implicit associations between social categories and nonhumans. *Psychological Science, 18*(2), 116–121.

Mahmud, H., Islam, A. K. M. N., Ahmed, S. I., & Smolander, K. (2022). What influences algorithmic decision-making? A systematic literature review on algorithm aversion. *Technological Forecasting and Social Change, 175*, 121390.

Ministry of Public Security of the People's Republic of China. (2020). https://www.mps.gov.cn/n2254098/n4904352/c7288071/content.html.

[中华人民共和国公安部. (2020). https://www.mps.gov.cn/n2254098/n4904352/c7288071/content.html.]

Morewedge, C. K. (2022). Preference for human, not algorithm aversion. *Trends in Cognitive Sciences, 26*(10), 824–826.

Newman, D. T., Fast, N. J., & Harmon, D. J. (2020). When eliminating bias isn't fair: Algorithmic reductionism and procedural justice in human resource decisions. *Organizational Behavior and Human Decision Processes, 160*, 149–167.

Niestadt, M., Debyser, A., Scordamaglia, D., & Pape, M. (2019). Artificial intelligence in transport: Current and future developments, opportunities and challenges. *European Parliamentary Research Service.*

Oleksy, T., Wnuk, A., Domaradzka, A., & Maison, D. (2023). What shapes our attitudes towards algorithms in urban governance? The role of perceived friendliness and controllability of the city, and human-algorithm cooperation. *Computers in Human Behavior, 142*, 107653.

Pribadi, U. (2021). Delivering Artificial Intelligence for Electronic Traffic Law Enforcement in Yogyakarta Region: Current Effort and Future Challenges. *In IOP Conference Series: Earth and Environmental Science, 717*(1), 012016.

Sinha, R. R., & Swearingen, K. (2001). Comparing recommendations made by online systems and friends. *DELOS, 106*.

Tayarani, M. (2020). Applications of artificial intelligence in battling against covid-19: A literature review. *Chaos, Solitons & Fractals.*

Tong, J. W. (2013). The harm and punishment of ten serious traffic violations are explained in detail. *Cars and Safety, 192*(12), 42–51.

[佟嘉旺. (2013). 详解十大严重交通违法行为的危害及处罚. *汽车与安全, 192*(12), 42–51.]

Trope Y, & Liberman N. (2003). Temporal construal. *Psychological Review, 110*(3), 403–421.

Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological Review, 117*(2),

440–463.
Vallacher, R. R., & Wegner, D. M. (1989). Levels of personal agency: Individual variation in action identification. *Journal of Personality and Social Psychology, 57*(4), 660–671.
Ward, A. F., Olsen, A. S., & Wegner, D. M. (2013). The harm-made mind: Observing victimization augments attribution of minds to vegetative patients, robots, and the dead. *Psychological Science, 24*(8), 1437–1445.
Wu, J. F., Yu, H. Y., Zhu, Y. M., Zhang, X. Y. (2020). Impact of Artificial Intelligence Recommendation on Consumers' Willingness to Adopt. *Journal of Management Science, 33*(5), 29–43.
[吴继飞, 于洪彦, 朱翊敏,张湘昀. (2020). 人工智能推荐对消费者采纳意愿的影响. *管理科学, 33*(5), 29–43.]
Yigitcanlar, T., & Cugurullo, F. (2020). The sustainability of artificial intelligence: An urbanistic viewpoint from the lens of smart and sustainable cities. *Sustainability, 12*(20), 8548.
You, S., Yang, C. L., & Li, X. (2022). Algorithmic versus Human Advice: Does Presenting Prediction Performance Matter for Algorithm Appreciation?. *Journal of Management Information Systems, 39*(2), 336–365.
Zhang, X. Y. (2021). Research on legal issues of road traffic electronic law enforcement. *Henan University.*
[张笑语. (2021). 道路交通电子执法法律问题研究(硕士学位论文). *河南大学*.]

The Effect of Algorithmic Monitoring on Compliance with Traffic Rules: From the Perspective of Construal Level

YIN jun[1], XIAO qinglu[1], WANG shiqi[1], CHEN jiakai[1], XU wanqiu[1], FAN lingxia[2]

([1]Department of Psychology, Ningbo University, Ningbo 315211)

([2]Department of Public Administration, Ningbo Administration Institution, Ningbo 315211)

[Abstract] As artificial intelligence continues to advance, algorithms find increasing applications across various domains, including education and transportation. In the realm of road traffic management, the escalating complexity of the traffic system poses challenges for traditional human monitoring, such as that carried out by traffic police. In light of this, there is a growing reliance on algorithmic monitoring, exemplified by electronic police systems. These systems offer extensive monitoring capabilities in terms of both time and space, providing an efficient means to uphold traffic order in the face of manual monitoring limitations. In order to enhance individuals' inclination and adherence to traffic rules, algorithmic monitoring serves as a compelling alternative to address the shortcomings of manual monitoring, which often involves blind spots and high operational costs. Observations from daily life experiences suggest that the broad coverage of algorithmic monitoring has a mitigating effect on traffic rule violations. Despite this intuitive understanding, there exists a notable gap in empirical research to substantiate these observations. While numerous studies have delved into the impact of algorithms, yielding findings related to both algorithm appreciation and aversion, there remains a need for a focused investigation into the specific influence of algorithmic monitoring on compliance with traffic rules and an exploration of the underlying mechanisms.

This research aims to address this gap by providing a nuanced understanding of how algorithmic monitoring shapes individuals' behavior in the context of obeying traffic rules. Drawing on construal-level theory, prior research has consistently shown that individuals tend to perceive humans at a high level of construal, while algorithms are typically construed at a low level. Moreover, traffic behaviors are also construed at different levels. Recognizing that the alignment of construal levels between the agent and behavior plays a pivotal role in influencing individuals, this paper posits a hypothesis: the effect of algorithmic monitoring on compliance with traffic rules hinges on the construal level of the traffic behavior, and the fit between the monitoring agent and monitored behavior acts as a mediating factor in this relationship. To address these considerations and test the proposed hypothesis, a preliminary investigation was conducted, selecting traffic behaviors with distinct construal levels (e.g., overspeed behavior as a low construal-level, and failure to give way to pedestrians as a high construal-level). In Study 1a, a situational test involving the description of various traffic events was employed to examine the influence of the monitoring

agent (human vs. algorithm) on the intention to comply with traffic rules across behaviors with different construal levels. Subsequently, Study 1b was conducted as a field study to assess the ecological validity of the findings from Study 1a. Building on these insights, Study 2 utilized a similar situational test as Study 1b but introduced error information indicating that the monitoring agent failed to detect traffic rule violations. This additional layer aimed to further investigate how the monitoring agent influences the intention to comply with traffic rules concerning behaviors with different construal levels.

We found that for traffic behaviors with low construal levels, compared to human monitoring, people had a stronger sense of fit with algorithmic monitoring, thereby enhancing their intention to comply with traffic rules. Conversely, for traffic behaviors with high construal levels, there was a stronger sense of fit under human monitoring, leading to a greater compliance intention with traffic rules. In summary, the monitoring agent influences individuals' intention to comply with traffic rules for behaviors at different construal levels, with the sense of fit playing an intermediate role. Further, after committing an error, the sense of fit induced by algorithmic monitoring decreased to a level comparable to human monitoring. Additionally, the positive effect on the intention to comply with traffic rules for behaviors with low construal levels disappeared. However, following an error in human monitoring, its monitoring effectiveness (i.e., compliance with traffic rules) for behaviors with high construal levels remained superior to that of algorithmic monitoring. Moreover, the mediating role of the sense of fit persisted. In essence, the monitoring effectiveness of algorithms is more significantly influenced by error information.

In summary, the enhancing effect of algorithmic monitoring on the intention to comply with traffic rules depends on the construal level of the observed traffic behavior, with the sense of fit playing a mediation role. Errors in algorithmic monitoring weaken its monitoring effectiveness for traffic behaviors with low construal levels. Hence, when the traffic management department chooses the monitoring agent, it should avoid indiscriminately using either algorithmic or human monitoring but consider the construal levels of traffic violations observed at the intersection.
[Keywords] algorithmic monitoring; human monitoring; compliance with traffic rules; construal level; algorithmic error